DOCUMENT RESUME

ED 255 579 TM 850 226

AUTHOR Chatman, Steven P.
TITLE The Relationship between Response Pattern Aberrance and Course Performance in Math Placement.
PUB DATE Mar 85
NOTE 38p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985). This paper is based, in part, on a doctoral dissertation, Texas A & M University, 1984.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Tests; College Freshmen; Higher Education; *Mathematics Achievement; *Predictive Validity; Regression (Statistics); *Responses; *Student Placement; Test Results
IDENTIFIERS *Extended Caution Index

ABSTRACT

University-level mathematics placement examinations might yield summary measures that are differentially accurate due to individual differences in the recall of previously-learned material. In particular, test scores in algebra and trigonometry may be subject to sporadic recall of solution procedures, resulting in aberrant response patterns. The potential of the individually-oriented standardized extended caution index (ECI) for contributing to prediction of future mathematics achievement was explored in two studies. The first study was designed to explore the predictive usefulness of the ECI as derived from pre-enrollment algebra and trigonometry exams. Specifically, analyses were designed to determine the ability of ECI to improve the prediction of performance in subsequent college-level mathematics courses. The results provided some evidence that supported the hypothesis of the sporadic recall, but only for trigonometry solution procedures and only for precalculus students. Therefore, a replication of the first study was undertaken to determine whether the marginally significant results were spurious. The results of Experiment 2 did not support the hypothesis of sporadic recall. Based on these two studies, there would appear to be no reason to routinely consider an incoming university student's placement test aberrance, as measured by the ECI, when determining a course placement. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED255579

THE RELATIONSHIP BETWEEN RESPONSE PATTERN ABERRANCE AND COURSE PERFORMANCE IN MATH PLACEMENT

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Steven P. Chatman

Measurement and Research Services
Texas A&M University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

S. P. Chatman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the 1985 Annual Meeting of the American Educational Research Association, Chicago. This paper is based, in part, on a doctoral dissertation (December, 1984, Texas A&M University). Requests for reprints should be sent to Steven P. Chatman, Measurement and Research Services, College Station, TX 77843.

INTRODUCTION

University-level mathematics placement examinations that are administered to incoming freshmen during pre-enrollment sessions might yield summary measures that are differentially accurate due to a lack of preparation by the majority of examinees and therefore, individual differences in the recall of previously learned material. A measure sensitive to this unwanted source of inaccuracy and which required no additional testing time could be very useful in institutional assessment for placement purposes. A class of response pattern indices that could potentially either correct for erroneous student characteristics or at least identify students whose responses did not fit an expectation of performance, are measures of response pattern aberrance. In this study, the predictive contribution of a member of this class of aberrance indices, the individually oriented standardized extended caution index ($ECI4_z$) (Tatsuoka, 1984), will be examined as simple and interactive regression terms and as a nominal marker for certain response patterns through analyses of predictive bias based on groups formed from the distribution of index values.

It has been hypothesized that examinee behaviors and characteristics can be meaningfully linked to measures of aberrance. Examples of examinee behaviors or examinee characteristics that could produce aberrant response patterns and therefore less appropriate test scores include: clerical recording errors, creative or over-analyzed responses, cheating (Hulin, Drasgow, & Parsons, 1983), guessing, sleeping, fumbling, plodding, cultural bias (Wright, 1977), a lack of preparation, sporadic study habits (Sato, 1975), unique background differences, differences in method of prior subject matter exposure, stress, motivational differences (Harnisch & Linn, 1981), and the consistent application of inappropriate solution strategies (Birenbaum & Tatsuoka 1983). Of these potential sources of aberrant response patterns, both a lack of preparation and guessing could be highly probable sources of individual aberrance in institutional mathematics assessment. However, a demonstrated lack of preparation that applies to the group and not only to certain individuals could tend to produce summary measures that reflect individual differences in recall.

Test scores from university mathematics placement testing are hypothesized to be subject to sporadic recall of solution procedures in pre-enrollment university mathematics (algebra and trigonometry). It is proposed that the recall of solution procedures in trigonometry, which are highly rule specific, are likely to be sporadic after an extended period of time since instruction. Sporadic recall as evidenced by an anomolous response pattern would, by definition, not follow group patterns since group-based expectations would be reflected in the index. It is further proposed that sporadic recall would be of less concern in algebra performance due to the more general solution strategies required and to their continued use in higher level mathematics. If sporadic recall, as measured by an anomolous response pattern, is a valid concern, then two students with the same total score, one with an expected scored response pattern and the other with an unexpected scored response pattern, are not of the same ability level. The unexpected pattern means that difficult items were answered correctly and easy

items were missed. It is hypothesized that the unexpected pattern reflects a higher actual level of ability which was inconsistently applied because it was inconsistently recalled. It is proposed that students who inconsistently solved test items, by missing easy items and answering difficult items, would subsequently outperform consistently responding students with the same total score since this type of aberrance would indicate a student with greater past ability. In other words, the aberrant students should differentially benefit from instruction since they possess a greater degree of acquired ability but are simply relearning specific solution procedures. A consistently responding student, on the other hand, must initially learn material measured by the items missed. This hypothesis assumes that students without knowledge of the material do not successfully mark correct answers. On the other hand, if guessing is a valid concern then the aberrance measures should have little bearing on subsequent performance. There is some evidence that aberrance measures are measures of guessing and this evidence will be presented subsequently.

The failure of students to study for the locally administered placement examinations is evidenced by a survey of the proceeding summer's entering students (Green, 1983). A sample of 667 incoming freshmen attending conferences in June and August of 1982 was asked to respond to a questionnaire concerning their motivation and performance on algebra and trigonometry tests administered during the conference. Ninety-four percent reported that they studied one hour or less for the algebra test and a similar majority (91%) studied one hour or less for the trigonometry test. Furthermore, 63% said it had been 2 or more years since they had taken an equivalent high school course in algebra and 53% said it had been 1 or more years since they had taken a trigonometry course. Given an extended period since completion of related subject matter and that little preparation for the examinations took place, it was likely that the students' retention of course content was sporadic.

Several researchers have proposed measures of the aberrance of individual scored response patterns in an effort to provide student performance information not represented in summary measures. Even a limited review of the available measures of response pattern aberrance will yield over 20 indices. The indices do have a common basis. Indices of aberrance are individually determined measures of the deviation of an observed pattern from an expected pattern. The indices differ in computational proceedures and in assumptions about response patterns. For a more complete description of group-based indices (indices based on summary measures) and various IRT-based indices the reader is referred to Cliff (1977), Harnisch and Linn (1981), Kane and Brennan (1980), Levine and Rubin (1979), Tatsuoka (1984), Tatsuoka and Linn (1983), Tatsuoka and Tatsuoka (1982), and Wright and Stone (1979). In this paper, the aberrance index employed was Tatsuoka's (1984) standardized extended caution index 4 ($EC14_z$) which uses an individually oriented, IRT expectation. The standardized extended caution indices are modifications and extensions of Sato's Caution Index (CI) (1975), as described in English by Tatsuoka and Tatsuoka (1980), that allow the direct comparison of standardized index values at different levels of ability.

There exists evidence to support some of the hypothesized applications of measures of aberrance in the areas of group differences, inappropriate solution algorithms, the identification of manipulated response patterns, and achievement measures. Supporting evidence is presented below.

Group Differences. Harnisch and Linn (1981) compared 10 indices based on observed response patterns for 6,300 randomly selected fourth, eighth, and eleventh grade Illinois public school children taking the Illinois Inventory of Educational Progress. Based on the intercorrelations between indices and the correlation of each index with total score, only the modified caution index (CI^*) was selected for further analysis. CI^* was able to identify regional and schooling differences among public schools in Illinois. CI^* was therefore useful as a measure of the applicability of a statewide achievement instrument to a state with various regional and local methods of education.

Inappropriate Solution Algorithms. Specifically, this research by K. K. Tatsuoka and associates, was based on signed-number arithmetic data where faulty algorithms that produced the types of consistent errors observed were constructed (Birenbaum & Tatsuoka, 1983). In a comparison of five extended caution indices, Tatsuoka and Linn (1983) reported that indices based on individual response patterns and theoretical person characteristic curves (ECI4 and ECI5) were effective in identifying persons consistently using inappropriate solution algorithms in the signed-number arithmetic problems.

Manipulated Response Patterns. Using random changes of correct or incorrect answers to simulate spuriously high and low scores, Rudner (1983) compared the selection efficiency of nine indices. He reported that most indices were able to identify high portions of examinees with spurious total scores and that spuriously high scores were easier to identify than spuriously low scores. Indices based on item response theory were consistently able to identify a higher portion of spurious scores than group-based indices. Further research based on tampered data, modified or generated to yield spurious scores, has consistently shown IRT-based indices to be capable of detecting spurious scores (Drasgow, 1982; Levine & Rubin, 1979; Levine & Drasgow, 1983).

Achievement Measures. Donlon and Fischer (1968) described the personal biserial index (r^*) and reported that the r^* supported the practice of reporting chance level SAT scores. Using 614 subjects with raw SAT scores within the chance range, they concluded that since a random response pattern would mathematically tend to produce zero or negative r^* values, then the mean r^* verbal biserial of 0.50 and the mean math biserial of 0.57 indicated that the results were "encouraging evidence for the possible use of ... [r^*] as a predictor of college success" (Donlon & Fischer, 1968, p. 112).

There is also evidence that various measures of response pattern aberrance used to examine group differences and to predict academic achievement are merely measures of guessing. This evidence is presented below.

Group Differences. Based on Harnisch and Linn's (1981) finding that variations in response pattern were related to school differences and Miller's 1982 (as described by Miller et al., 1984) finding that these variations could be found at the class level, Miller et al. (1984) attempted to apply CI^* to the eighth grade sample of the International Association for the Evaluation of Educational Achievement study of curricula, classroom processes, and student outcomes. They focused only on the eighth grade United States sample (6,788 students from 298 classes), and were unable to find any relationship between CI^* and instructional variation or textbook used. Their conclusion was that CI^* was nothing more than a measure of guessing and provided no additional information beyond total test score.

Prediction of Achievement. In the first examination of the contribution of indices to the prediction of academic achievement, college grade point average, Fischer (1970) reported that r^* did not significantly contribute to concurrent validity. Hulin, Drasgow, and Parsons (1983) suggest that r^* should have been viewed as a moderator variable and that inclusion of r^* interaction terms with SAT verbal and math scores might have proved useful.

More recently, Smilaycoff and Blixt (1984) examined the relationship between CI^*, derived from the administration of an elementary algebra skills placement examination to 229 university students admitted with conditional status, and course grades in algebra and college algebra separately for higher ability and lower ability students. They reported that the negative regression coefficient for CI^* was significantly different from zero ($p<.05$) for low ability students (total score less than 28 on the 35 item test), but not for higher ability students. They concluded that CI^* was related to guessing for lower ability students and the guessing tended to produce erroneously high total test scores for these lower ability students with higher CI^*s.

Experiment 1

Purpose of the Study

This study was designed to explore the predictive usefulness of the individual-based standardized extended caution index ($ECI4_z$) (Tatsuoka & Tatsuoka, 1982) as derived from pre-enrollment algebra and trigonometry achievement examinations. Specifically, analyses were designed to determine the ability of $ECI4_z$ to improve the prediction of performance in subsequent college-level mathematics courses.

Two methods of analysis which directly examine the contribution of standardized extended caution indices to the prediction of subsequent performance are linear regression, treating the caution indices as main and interactive terms, and the comparison of regression slopes for predictive bias for aberrantly versus consistently responding groups.

Consistent with the hypothesis of sporadic recall leading to differential benefit from subsequent instruction, if the standardized extended caution index ($ECI4_z$) contributes linear predictive

information, in the direction hypothesized and beyond that contributed by total score, the coefficients associated with the indices as main and/or interactive effects should be significant and should be positive. The coefficients should be positive since positive indices suggest aberrance.

To further address the research hypothesis and to couch "aberrance" in a way that could be easily communicated to advisors (i.e., by "flagging" performance records of some students as probable underestimates of ability), groups formed on the basis of the distribution of $ECI4_z$ were considered as nominal categories for an examination of predictive bias. If the relationship between test score and subsequent performance was such that the subsequent performance of members of the aberrantly responding group would be systematically underpredicted by use of either a common regression or the regression based on the consistently responding group, then the hypothesis of sporadic recall would be supported.

Experiment 1

Method

Subjects and Course Placements

Test and item characteristics for both algebra examinations and the trigonometry examination used in this study were based on a sample of 710 students admitted to the university for the Fall semester of 1983. These students attended a mid-August conference required of all regularly admitted freshmen. At this conference, all students took one of two algebra achievement examinations and those with previous exposure to trigonometry (N=536) took a trigonometry achievement test. The students attending this conference were apparently representative of the mathematics ability of the entire 1983 fall freshman class since their Scholastic Aptitude Test mathematics average was 552, as compared to the class average of 555 (N=5,209), and their College Board Math Achievement Level I test average was 527, which was equal to the fall, 1983 freshman class (N=4,625). The group variances were also highly similar to freshman class values. The standard deviation of the Scholastic Aptitude Test mathematics scores was 87, compared to a freshman class standard deviation of 90, and their Mathematics Level I Achievement Test standard deviation variance was 86, same as the freshman class standard deviation of 86.

After meeting with advisors, the students registered for their first semester courses. The research sample was limited at course conclusion by schedule changes and the availability of semester point totals within course sections. In addition, students with scores of zero or 40 on the 40 item tests were excluded since their abilities can not be estimated by the maximum likelihood method. The semester point totals were supplied by mathematics faculty members and were converted to standardized scores within course sections. The standardized semester point totals constitute the performance criterion.

Standardized criterion measures were available for 270 precalculus students, only 209 with trigonometry ability estimates, and 144 calculus students, 138 with trigonometry ability estimates.

Instrumentation

The examinations were from the locally designed precalculus skills tests (Algebra Form A, Algebra Form B, and Trigonometry). All examinations were 40 item, 5 option, mulitiple-choice tests with 45 minute time limits. The exams were designed by a mathematics faculty member to measure skills and knowledge consistent with the algebra objectives for precalculus courses. The Trigonometry examination was similarly constructed and was consistent with the trigonometry objectives for precalculus courses. The internal consistency reliabilities ($\underline{K\text{-}R}_{20}$) were 0.88 ($\underline{N}$=366) for Algebra Form A, 0.90 ($\underline{N}$=355) for Algebra Form B, and 0.83 ($\underline{N}$=536) for Trigonometry. For all administrations, students were instructed that there was no penalty for missed responses. Average scores and the standard deviations of scores on the Algebra Form A and Form B examinations were 17.46 and 7.73, and 16.81 and 8.27, respectively.

The algebra tests were parallel forms in the sense that they were designed, item by item, to be parallel and that the means, standard errors of measurement, and standard deviations were not significantly different. The $\underline{t}$ associated with mean examination differences was 1.09 which was not significant at the .1 level. The $\underline{F}$ value for comparing variances and standard errors were 1.07 ($\underline{p}$<.74) and 1.03 ($\underline{p}$<.60).

All analyses used the common measures and standards that follow. The one-parameter ability estimates were used instead of total score. The analyses were conducted at two levels, precalculus and calculus, since there was no logical reason to equate performance between these different levels of content and ability. Criterion performance was defined as a standard score based on student semester point totals within course sections to control for grading differences. The Form A and Form B algebra test results were treated as equivalent measures. Lastly, those analyses which included trigonometry measures excluded students who did not submit trigonometry scores. Students with perfect or zero scores on one examination were excluded from analyses using the results of that examination since ability estimates are not defined for zero or perfect scores in maximum likelihood estimation.

Computer Program. The computation of indices and the Rasch person and item estimates were performed using SECI (Chatman & Nelson, 1984), a SAS Matrix procedure program. The program program includes Andersen's (1973) correction for biased estimates of item difficulties (as described by Wright & Stone [1979]). The Rasch calibration method used was an iterative estimation procedure based on the algorithm and the maximum likelihood routine suggested by Wright and Stone (1979, pp. 64-65). The program computes the initial estimates for maximum likelihood solution iteration by incorporating the computation algorithms suggested by Wright and Stone (1979, pp. 61-62). The criterion for convergence was a change of less than 0.00001 in the estimates between sequential

iterations. The program's accuracy was verified using the examples provided in Wright and Stone (1979). $ECI4_z$s were computed using the formula suggested by Tatsuoka (1984, p. 105).

$$ECI4_{zi} = \frac{\sum_{j=1}^{n} P_{ij}(P_{ij} - T_i)}{\left(\sum_{j=1}^{n} s^2_{ij}(P_{ij} - T_i)^2\right)^{1/2}} - \frac{\sum_{j=1}^{n} y_{ij}(P_{ij} - T_i)}{\left(\sum_{j=1}^{n} s^2_{ij}(P_{ij} - T_i)^2\right)^{1/2}} \quad (12)$$

(Tatsuoka, 1984, p. 105).

In these equations:

y_{ij} = observed binary responses;
P_{ij} = elements of the probability matrix;
G_j = average probability of responses by items j=1, ...,n;
T_i = average probability of responses by persons i=1, ..., N;
and, s^2 = $P_{ij}(1 - P_{ij})$.

Results

Descriptive Statistics and Zero-order Correlations

Standard descriptive statistics for the major variables are presented in Table 1. To exclude the influence of extreme cases, only students with aberrance indices between 3.0 and -3.0 were considered.

Predictably, students placed into precalculus courses had lower mean algebra ability (-0.802) and trigonometry ability (-1.149) than students placed in to calculus courses (mean algebra ability = 0.494; mean trigonometry ability = -0.274). The mean algebra and trigonometry abilities for all students indicate that the distributions were positively skewed. There was also a tendency for $ECI4_z$ to be greater for precalculus students. In other words, the students placed into calculus courses tended to exhibit more consistent response patterns. This suggests that total score might be negatively correlated with index values.

INSERT TABLE 1 HERE

The relationship between total score and index values is examined in Table 2. Zero-order correlations using all students are presented in Table 2. Tables 3 and 4 present the zero-order correlations for precalulus students and calculus students, respectively.

An examination of correlations presented in Table 2 shows that index values for algebra measures were uncorrelated with algebra ability

and index values for trigonometry measures were only slightly negatively correlated with trigonometry ability. Even for the highest correlation (r = -0.156), which was the correlation between $ECI4_z$ and trigonometry ability, the index value accounts for less than 2.5 percent of the variance in trigonometry ability.

INSERT TABLES 2, 3, & 4 HERE

Predictive Validity of Indices

To determine whether or not $ECI4_z$ should be considered when predicting student performance in subsequent mathematics courses, the two $ECI4_z$s, one based on the observed response pattern for the algebra examination and the other based on the observed response pattern for the trigonometry examination, wer considered as main and interactive effects only after performance on the algebra and trigonometry examinations was considered. In linear regression, performance on the two examinations was controlled by sequentially entering the algebra ability estimate and the trigonometry ability estimate as the first two dependent measures. The contributions of main and interactive terms can then be determined by F tests of the type I sequential sums of squares. The results of full model sequential sums of squares are presented as Table 5 for precalculus students and Table 6 for calculus students.

INSERT TABLES 5 & 6 HERE

Addressing Hulin, Drasgow, and Parsons's (1983) concern for the possibility of an interaction between test performance and aberrance values, there was no evidence of an interactive effect between $ECI4_z$ and test performance for precalculus or calculus students. There was evidence to justify the use of $ECI4_z$ values in the prediction of course performance. However, that evidence was only relevant to performance in precalculus courses.

The regression weights associated with each term for a precalculus course performance prediction model including algebra and trigonometry ability estimates and aberrance indices appear in parentheses in the following equation:

$$y = 0.415 + a(0.299) + b(0.322) - ECI4_{za}(0.098) - ECI4_{zb}(0.137); \quad (13)$$

where y = predicted performance;
a = algebra ability;
b = trigonometry ability;
$ECI4_{za}$ = $ECI4_z$ for algebra pattern;
and, $ECI4_{zb}$ = $ECI4_z$ for trigonometry pattern.

This model accounted for 19.4% of course performance variance (n=209). Algebra and trigonometry abilities alone accounted for about 13.8% of precalculus course performance variance. When trigonometry measures were excluded in an additional precalculus analyses the porportion of variance accounted for was .118 (n=271), the intercept dropped to 0.143,

the algebra ability coefficient increased to 0.428, and the $ECI4_{za}$ coefficient remained largely unchanged at 0.132. Algebra ability alone accounted for 12.7% of the variance in precalculus course performance.

Examinations of Predictive Bias

The second research question was designed to address predictive bias using both the method of predictive bias analysis suggested by Jensen (1980, pp. 379 & 455-457) and Johnson-Neyman predictive bias analyses (Rogosa, 1981). The first method of determining test bias defines bias as the presence of statistically significant differences between the nominal groups in the slope, intercept, or standard error of estimate of the regression of test score on criterion performance for the nominal groups. More simply, is the relationship between the test's measure of ability and the criterion measure of interest different for nominally classified groups? The Johnson-Neyman techniques for population within-group regression as described by Rogosa (1981) will consist of regression slope comparisons and the simultaneous regions of significance for comparisons with unequal slopes.

The nominal classifications used in these analyses were aberrantly responding and consistently responding. The classifications were made on the basis of $ECI4_z$. To insure that an adequate number of students would be placed in the aberrantly responding group and maintain some validity of classification, group membership was made relative to an $ECI4_z$ value of 0.4195 for algebra group classification and 0.4373 for trigonometry group classification. These values defined the actual top one-third of the two distributions. In Experiment 2, the standard score value of 0.43 was used.

Nominal group classification allowed the comparison of group mean differences. The comparisons used were based on analysis of variance F tests although t tests would have also been appropriate for two group comparisons. The results of group comparisons are presented in Table 7.

There were two significant mean differences found. The algebra abilities of students with aberrant algebra patterns differed from those with consistent algebra patterns for calculus students (F=5.02, p<.0265). The aberrant algebra group of calculus students had higher mean algebra ability (0.746) than the consistent algebra group of calculus students (0.417). The other significant difference was found between the aberrant and consistent trigonometry groups of precalculus students (F=7.10, p<.0083) on course performance. In this comparison, the aberrantly responding group had a significantly lower mean course performance (-0.283) than the consistently responding group (0.098).

INSERT TABLE 7 HERE

The relationships between mathematical abilities and course performance for aberrant and consistent groups are presented in Table 8.

Regression Slopes. The results of examinations of regression slope comparisons as described by Jensen (1980, pp. 455-457) are presented in

Table 9. Regression lines of trigonometry ability on course performance by aberrance group were found have significantly different slopes for students in precalulus courses (t=2.03, n=209) but not for students in calculus courses (t=0.05, n=238) or for algebra scores in either group ([t=0.339, n=270 for precalulus] and [t=0.030, n=138 for calculus]). For precalculus students, the beta estimate of 0.673 for the aberrant trigonometry group was significantly higher than the beta estimate of 0.241 for the consistent trigonometry group. The Johnson-Neyman tests of parallel regressions (Rogosa, 1981) yielded equivalent results with only the trigonometry examination exhibiting bias for the precalculus group (F =4.1099, p=.0439).

Standard Error of Estimates. The second criterion of test bias, significantly different standard error of estimates, was examined and the results are presented in Table 10. Two standard error of estimates comparisons were found be significantly different. The standard error of estimates for the aberrant precalculus algebra group and for the aberrant calculus trigonometry group were significantly higher than the respective consistent groups.

INSERT TABLES 8, 9, & 10 HERE

Intercepts. As a final test of predictive bias, the intercepts were examined for significant differences. According to Jensen (1980, p. 456), the proper test of intercept differences is an examination of adjusted group examination mean differences where performance on the criterion functions as a covariate. These comparisons yielded no significantly different adjusted ability estimates and therefore no significantly different intercepts.

Simultaneous Regions of Significance. The Johnson-Neyman simultaneous regions of significance, as described by Rogosa (1981), were computed for the case of a significant interaction. The only comparison yielding a significant interaction between nominal group membership and ability when predicting course performance occurred with the trigonometry examination for precalculus students. Using the computational method described by Willson (1984) for the single predictor case, a .05 confidence band describing the regions of abilities for which prediction was significantly different was constructed. This Johnson-Neyman technique consists of defining two points along the ability axis beyond which abilities can be considered differently predicted. The region for which there was no statistically significant difference was from trigonometry ability -3.113 to ability 0.097 (df= 2, 261). Given an evidenced range of abilities between about -3 and 2, the region of interest was abilities greater than 0.097 and less than 2. These were students with greater than average trigonometry abilities who were placed into precalculus. For these students, use of a common regression or the major group regression would systematically underpredict subsequent course performance.

Discussion

Contribution of Indices to Prediction

To determine the contribution made by aberrance indices beyond total score measures, the additional contribution of $ECI4_z$ was determined through linear regression where the trigonometry- and algebra-based ability estimates entered solution after variance was accounted for by algebra and trigonometry ability. For precalculus students, both $ECI4_z$ for algebra ($p<.0244$) and $ECI4_z$ for trigonometry ($p<.0208$) were significant at the .05 level of significance. The indices were marginally useful in predicting calculus performance and there was no evidence of an interaction between $ECI4_z$ and ability as suggested by Hulin, Drasgow, and Parsons (1983, p. 118-121). The regression coefficients associated with the two $ECI4_z$ terms were negative. This means that slight increases in predicted performance would be made for students with very consistent response patterns since they have negative index values. Predicted performance, as predicted from algebra and trigonometry abilities, would be slightly adjusted downward for students with aberrant response patterns since their index values were positive. According to the hypothesis of sporadic recall, the regression weights should have been positive. The relationships between regression coefficients and course performance indicated that prediction might be different for aberrantly and consistently responding students even though aberrance and ability were not significantly correlated.

The contribution made by the two $ECI4_z$s was significant only for precalculus students. Since the inclusion of the two $ECI4_z$ indices increased the percent of variance explained by less than 5% for precalculus students and were not significant for calculus students, the indices probably should not be standardly used as predictors in applied settings.

A second type of analysis was conducted to determine whether or not the relationships between test performance and course performance were equivalent for aberrantly and consistently responding groups. To perform these analyses of test bias, students were nominally defined as aberrant or consistent based on the distribution of $ECI4_z$ values for all students. Operationally, students with aberrance indices in the top one-third of the distribution were nominally categorized as aberrant and students with indices in the remaining two-thirds of the distribution were categorized as consistent.

Predictive Bias as a Function of Aberrance Classification

Since there were two measures of mathematics ability and two course placements, there were four bias comparisons. Test bias was indicated in three of the four comparisons. The standard error of estimates for aberrant precalculus students, based on the algebra examination, and for aberrant calculus students, based the trigonometry examination, were significantly higher than their respective consistent groups. Given

that the test should be more accurate for students who respond in a consistent manner, these results were anticipated. The third case of test bias occurred for precalculus students on the trigonometry test. This bias was nonparallel regression slopes for the aberrant and consistent groups. The slope of test regression on course performance was steeper for aberrantly responding students than for consistently responding students. A Johnson-Neyman simultaneous regions of significance analysis yielded two regions of trigonometry ability for which course performance predictions were significantly different. The lefthand region was abilities less than -3.11, which falls outside of the region of interest since ability estimates of this nature were excluded from consideration as outlying cases. The righthand region of significant difference was abilities greater than 0.097 and was within the region of interest.

Use of a common regression or the major group regression would therefore result in a systematic underprediction of precalculus course performance for aberrantly responding students with greater than average trigonometry abilities. Regression intercepts were not significantly different in any of the comparisons.

Based on these results, it was concluded that the trigonometry test items required a specificity of recollection that was not required by the algebra test and that the test scores for more able aberrant students were probably underestimates of their true abilities.

Experiment 2

Purpose of this Study

While the results of this first study indicated that $\underline{ECI4}_z$ might provide useful information for prediction of precalculus students' performance, the same was not true for calculus students. There might be two reasons for this lack of usefulness. First, as suggested by Tatsuoka and Tatsuoka (1980), aberrance may be of less concern for higher ability students. Second, the calculus course requirements were not as closely aligned with the content of the algebra and trigonometry examinations. The algebra and trigonometry examinations were specifically designed to reflect the precalculus course requirements.

The results suggested that standardized indices of aberrance might provide useful information, beyond that found in achievement total score, in the prediction of performance for some students. For aberrantly responding students, there was a range of abilities in which ability was systematically underpredicted by the common regression or the major group regression. These students were of greater than average ability and responded in an aberrant fashion.

The interpretation suggested, that the underprediction was due to sporadic recall of the trigonometry solution procedures, suggests an interesting, although limited area of future research. The results indicated that ability, aberrance, and examination content can interact to yield inaccurate estimates of ability. The results have some

intuitive appeal. When students are assessed in a content area requiring highly specific recall and the students do not prepare for the examination, it seems reasonable that higher ability students might be able to correctly answer some difficult items but fail to answer some easier items. These students' total scores would then underrepresent their true abilities. Given this interpretation, tudents with aberrant patterns and greater than average ability would differentially benefit from preparation or subsequent exposure.

The most obvious research suggested by these results is an attempted replication. It is possible that the marginally significant results found in Experiment 1 were spurious. Therefore, the analyses used in the first phase of the study were replicated in the second phase. The second experiment was based on students attending conferences during the following summer. The major changes between the replication analyses and the initial analyses were 1) the use of end of semester grade in math course instead of a standardized score based on semester point totals within course sections, 2) the much larger samples used in Experiment 2 (e.g., n=557 for course performance in Experiment 1 vs n=4579 for course grade in Experiment 2), and 3) the use of various Experiment 1 test form item sorts in Experiment 2. Test form items were sorted in Experiment 2 to reduce the likelihood of shared responses given the testing conditions used in Experiment 2. Given that Experiment 2 was an attempted replication, only the findings that were inconsistent with those found in Experiment 1 will be discussed. All results from Experiment 2 can be found in Tables 11 through 20.

Results

Descriptive variables and zero-order correlations were consistent from Experiment 1 to Experiment 2. An examination of the type 1 sequential sums of squares from Experiment 2 showed some inconsistency between Experiments 1 and 2 for precalculus students. In Experiment 1, trigonometry ability and the associated $ECI4_z$ were both significant main effects ($p<0.05$). In Experiment 2, neither term was significant at the 0.05 level although the measured probabilities of 0.06 and 0.08 indicate a similarity between the two regressions.

Analysis of variance group comparisons from Experiment 2 produced mixed support for the results from Experiment 1. On dependant variable algebra ability: 1) the aberrant and consistent precalculus groups did not differ in Experiment 1 but did differ in Experiment 2; and, 2) the calculus groups were different in Experiment 1 but were not different in Experiment 2. On dependent variable trigonometry ability the precalculus groups did not differ in Experiment 1 but did differ in Experiment 2. On dependent variable course performance the algebra-based precalculus groups did not differ in Experiment 1 but were different in Experiment 2. One consistent difference was found. In both experiments the trigonometry-based precalculus groups differed in course performance with the consistent group outperforming the inconsistent group. However, the groups trigonometry scores did not differ in Experiment 1 but do differ in Experiment 2.

Standard error of estimate predictive bias comparisons in Experiment 2 showed mixed support for the results from Experiment 1. In both cases, the algebra to precalculus simple regressions were different for precalcus groups, with a larger error associated with the aberrant groups. The algebra to course performance regressions were similarly less accurate for the calculus students in Experiment 2 although there was no difference in Experiment 1. Likewise, the trigonometry to course performance regression for calculus students was significant in Experiment 1 but not in Experiment 2. The trigonometry regression slope difference found for precalculus students in Experiment 1 was not supported in Experiment 2.

DISCUSSION

Consistent from Experiment 1 to Experiment 2, the predictive contributions of $\underline{ECI4}_z$s as main and interactive effects was moderate at best and the weights associated with significant $\underline{ECI4}_z$s were negative. The negative weights do not support the hypothesis of sporadic recall since negative aberrance measures indicate consistent response patterns and therefore an additive contribution to predicted course performance. Likewise, aberrant response patterns are positive and the use of the negative weights would result in a lesser predicted course performance for aberrantly responding students.

In the first experiment, there was some evidence that supported the hypothesis of the sporadic recall, but only for trigonometry solution procedures and only for precalculus students. Specifically, the course performance of some members of the aberrantly responding precalculus group was underpredicted by the major group regressions. This result was not supported in Experiment 2. Beyond simply attributing the replication failure to some data characteristic in Experiment 1, there are two alternative explainations that bear examination. First, the aberrant groups in Experiment 1 were formed at the top one-third level to protect group sizes and that is not a concern in Experiment 2. Second, the subjects in Experiment 1 attended an August conference and the subjects in Experiment 2 included students attending conferences in late May, June, July, and August.

If the designation of aberrance is restricted to the top one-tenth of the $\underline{ECI4}_z$ in Experiment 2, the correlations between math ability and course performance show no indication of supporting the results from Experiment 1. The correlation between algebra ability and course performance for the aberrant precalculus group is 0.338 ($\underline{n}$=289) vs 0.375 ($\underline{n}$=2009) for the consistent precalculus group. The correlation between trigonometry ability and course performance for the aberrant precalculus group is 0.115 ($\underline{n}$=293) vs 0.193 ($\underline{n}$=1680) for the consistent precalculus group. Similar results were found for calculus groups. If the replication were restricted to students attending conferences in August of the second summer and a one-third cut-off were used, the group sizes prevent any comparisons for calculus students and the comparisons for precalculus students do not support the results of Experiment 1 (e.g., the correlation between trigonometry ability and course performance for the aberrant precalculus group was 0.276 ($\underline{n}$=81) vs 0.224 ($\underline{n}$=167) for the consistent group).

Based on these results, there would appear to be no reason to routinely consider an incoming univerisity student's placement test aberrance, as measured by $\underline{EGI4}_2$, when determining a course placement. The marginal contribution made by this index would be offset by the difficulty of explaining the index to the students' advisors.

REFERENCES

Birenbaum, M., & Tatsuoka, K.K. (1983). The effect of a scoring system based on the algorithm underlying the students' response patterns on the dimensionality of achievement test data of the problem solving type. Journal of Educational Measurement, 20(1), 17-21.

Chatman, S.P., & Nelson, R.B. (1984). SPCI [Computer program]. College Station, TX: Texas A&M University, Measurement and Research Services.

Cliff, N. (1977). A theory of consistency of ordering generalizable to tailored testing. Psychometrika, 42(3), 375-399.

Donlon, T.F., & Fischer, F.E. (1968). An index of an individual's agreement with group-determined item difficulties, Educational & Psychological Measurement, 28, 105-113.

Drasgow, F. (1982). Choice of test model for appropriateness measurement. Applied Psychological Measurement, 6(3), 297-308.

Fischer, F. (1970). Some properties of the personal biserial index. Journal of Educational Measurement, 7(4), 275-277.

Green, J.C. (1983). New freshmen evaluate the TAMU mathematics placement tests (Tech. Rep. No. 2). College Station, TX: Texas A&M University, Measurement & Research Services.

Harnisch, D.L. (1983). Item response patterns: Applications for educational practice. Journal of Educational Measurement, 20(2), 191-206.

Harnisch, D.L., & Linn, R.L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. Journal of Educational Measurement, 18(3), 133-146.

Hulin, C.L., Drasgow, F., & Parsons, C.K. (1983). Appropriateness measurement. In C.L. Hulin, F. Drasgow, & C.K. Parsons (Eds.), Item response theory (pp. 110-151), Homewood, IL.: Dow Jones-Irwin.

Jensen, A. (1980). Bias in mental testing (pp. 455-456). New York, NY: The Free Press.

Kane, M.T., & Brennan, R.L. (1980). Agreement coefficients as indices of dependability for domain-referenced tests. Applied Psychological Measurement, 4(1), 105-126.

Levine, M.V., & Drasgow, F. (1983). Appropriateness measurement: Validating studies and variable ability models. In D.J. Weiss (Ed.), New horizons in testing: Latent trait test theory and computerized adaptive testing (pp. 110-134). New York, NY: Academic Press.

Levine, M.V., & Rubin, D.F. (1979). Measuring the appropriateness of multiple-choice test scores. Journal of Educational Statistics, 4, 269-290.

Miller, M.D. (1982). Measuring between-groups differences in instruction. Unpublished doctoral dissertation, University of California, Los Angeles.

Miller, M.D., Linn, R.L., & Harnisch, D.L. (1984). S-P analysis of eighth grade mathematics achievement. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Rogosa, D. (1981). On the relationship between the Johnson-Neyman region of significance and statistical tests of parallel within-group regressions. Educational and Psychological Measurement, 41, 73-84.

Rudner, L.M. (1983). Individual assessment accuracy. Journal of Educational Measurement, 20(3), 207-219.

Sato, T. (1975). The construction and interpretation of S-P tables. Tokyo: Meiji Tosho. (in Japanese).

Smilaycoff, M.P., & Blixt, S.L. (1984). An investigation of the modified caution index in the prediction of college success. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Tatsuoka, K.K. (1984). Caution indices based on item response theory. Psychometrika, 49(1), 95-110.

Tatsuoka, K.K., & Linn, R.L. (1983). Indices for detecting unusual patterns: Links between two general approaches and potential applications. Applied Psychological Measurement, 7(1), 81-96.

Tatsuoka, K.K., & Tatsuoka, M.M. (1980). Detection of aberrant response patterns and their effect on dimensionality (Res. Rep. 80-4-ONR). Urbana, IL: University of Illinois, Computer-based Education Research Laboratory.

Tatsuoka, K.K., & Tatsuoka, M.M. (1982). Standardized extended caution indices and comparisons of their rule detection rates (Res. Rep. 82-4-ONR). Urbana, IL: University of Illinois, Computer-based Education Research Laboratory.

Willson, V.W. (in press). Analysis of interactions in ATI research. In C. Reynolds & V. Willson (Eds.), Methodological and statistical advances in the study of individual differences. New York, NY: Plenum.

Wright, B.D. (1977). Solving measurement problems with the Rasch model. Journal of Educational Measurement, 14(2), 97-115.

Wright,B.D., & Stone, M.H. (1979). Best Test Design. Chicago, IL: MESA Press.

APPENDIX A

SECI

Note: Necessary system cards should be included to define the location of the input data file, the output file, and to execute SAS library.

```
*-------------------------------------------------------------------*
*    THIS PROC MATRIX ROUTINE PERFORMS RASCH ITEM CALIBRATION       *
*    AND GOES ON TO COMPUTE TO CAUTION INDICES                      *
*                                                                   *
*    BY STEVE CHATMAN AND RANDY NELSON                              *
*                                                                   *
*-------------------------------------------------------------------*;
DATA A; INFILE DATAIN;
  INPUT @15 (X1-X40) (1.);
DATA B;INFILE DATAIN;
  INPUT @1 ID /.;
PROC MATRIX;
  FETCH YIJ DATA=A;          *NAME THE SCORE MATRIX YIJ;
  FETCH NAME DATA=B;
  ITEMS=NCOL(YIJ);           *NUMBER OF ITEMS ON TEST;
  PERSONS=NROW(YIJ);         *NUMBER OF TESTS;
*-------------------------------------------------------------------*
*                      SCORE THE TEST                               *
*            THIS PROGRAM EXPECTS THE ANSWER KEY                    *
*              TO BE THE FIRST RECORD ENTERED                       *
*-------------------------------------------------------------------*;
  DO I=1 TO ITEMS;
  DO J=2 TO PERSONS;
  IF YIJ(1,I)=YIJ(J,I) THEN YIJ(J,I)=1;
  ELSE YIJ(J,I)=0;           *CHANGE YIJ FROM A MATRIX OF RESPONSES;
  END; END;                  *TO A MATRIX OF ITEM SCORES;
  YIJ=YIJ(2:PERSONS,1:ITEMS);    *ELIMINATE THE KEY FROM THE MATRIX;
  NAME=NAME(2:PERSONS,1);        *ELIMINATE THE KEY FROM THE MATRIX;
*-------------------------------------------------------------------*
*             MODIFY THE ITEM SCORE MATRIX TO                       *
*           ELIMINATE USELESS ITEMS AND PERSONS                     *
*         SEE WRIGHT AND STONE, BEST TEST DESIGN                    *
*-------------------------------------------------------------------*;
BEGIN:
  ISCORE=YIJ(+,);            *TOTAL NUMBER OF CORRECT RESPONSES/ITEM;
  NPERSONS=NROW(YIJ);        *NUMBER OF PERSONS IN THE MATRIX;
  NITEMS=NCOL(YIJ);          *NUMBER OF ITEMS IN THE MATRIX;
  V1=LOC(ISCORE NPERSONS&ISCORE 0);  *LOCATE ITEMS WITH PERFECT OR;
  YIJ=YIJ(1:NPERSONS,V1);            * ZERO SCORES AND DELETE THEM;
  PSCORE=YIJ(,+);            *RAW SCORE FOR EACH PERSON;
  NITEMS1=NCOL(YIJ);         *NUMBER OF ITEMS IN THE MODIFIED MATRIX;
  V2=LOC(PSCORE 0&PSCORE NITEMS1);   *LOCATE PERSONS WITH PERFECT OR;
  YIJ=YIJ(V2,1:NITEMS1);             * ZERO SCORES AND DELETE THEM;
```

```
  NAME=NAME(V2,1);                    * INCLUDING THE NAME MATRIX;
  NPERSON1=NROW(YIJ);       *NUMBER OF PERSONS IN THE NEW MATRIX;
  IF NPERSON1=NPERSONS & NITEMS=NITEMS1 THEN GOTO GOODATA;
  ELSE GOTO BEGIN;          *IF MODIFICATION IS COMPLETE, PROCEED;
*-------------------------------------------------------------------*
*                PERFORM RASCH ITEM CALIBRATION                     *
*                    USING THE PROX METHOD                          *
*         SEE WRIGHT AND STONE, BEST TEST DESIGN                    *
*-------------------------------------------------------------------*;

GOODATA: NOTE 'NUMBER OF ITEMS USED FOR CALIBRATION'; PRINT NITEMS;
  NOTE 'NUMBER OF ITEMS DELETED';
  ITEMSDEL=ITEMS-NITEMS; PRINT ITEMSDEL;
  NOTE 'NUMBER OF PERSONS USED IN CALIBRATION'; PRINT NPERSONS;
  NOTE 'NUMBER OF PERSONS DELETED';
  PERSDEL=PERSONS-NPERSONS; PRINT PERSDEL;
  R=1:NITEMS-1;              *CREATE R: VECTOR OF ITEMS CORRECT;
  N=J. (NITEMS-1,1,0);       *INITIALIZE N: VECTOR OF SCORE GROUP SIZE;
  DO I=1 TO NPERSONS;                *COUNT THE PERSONS RECEIVING EACH;
  N(PSCORE(I,),)=N(PSCORE(I,),)+1; *    TOTAL SCORE AND STORE IN N  ;
  END;
  XI=LOG((NPERSONS-ISCORE)#/ISCORE);  *INITIAL RASCH ITEM VALUES;
  XDOT=(SUM(XI))#/NITEMS;             *MEAN OF INITIAL ITEM VALUES;
  YR=LOG(R#/(NITEMS-R));              *INITIAL RASCH PERSON VALUES;
  NYR=N'#YR;
  YDOT=(SUM(NYR))#/NPERSONS;          *MEAN OF INITIAL PERSON VALUES;
  D=SUM((XI-XDOT)##2)#/(2.89#(NITEMS-1));          ******************;
  BEE=SUM(N'#((YR-YDOT)##2))#/(2.89#(NPERSONS-1));* D BEE AND D ARE*;
  G=BEE#D;                                         *   PARTS OF THE *;
  X=SQRT((1+D)#/(1-G));                            *EXPANSION FACTOR*;
  Y=SQRT((1+BEE(#/(1-G));                          ******************;
  B=Y#(XI-XDOT);                      *PROX ESTIMATE OF ITEM VALUES;
  NOTE 'INITIAL ITEM VALUES'; PRINT B;
  SED=Y#(SQRT(NPERSONS*/(ISCORE#(NPERSONS-ISCORE))));
  NOTE 'STANDARD ERRORS FOR ITEM VALUES'; PRINT SED;
  THETA1=X#YR;                       *PROX ESTIMATE OF PERSON VALUES;
  NOTE 'INITIAL PERSON VALUES'; PRINT THETA1;
  SEB=X#(SQRT(NITEMS#/(R#(NITEMS-R))));
  NOTE 'STANDARD ERRORS FOR PERSON VALUES'; PRINT SEB;
*-------------------------------------------------------------------*
*                PERFORM RASCH ITEM CALIBRATION                     *
*                BY MAXIMUM LIKELIHOOD ROUTINE                      *
*           SEE WRIGHT AND STONE, BEST TEST DESIGN                  *
*-------------------------------------------------------------------*;
  ITER=0;
ITER8; SUMD2=0; RPROB=J.(NITEMS-1,NITEMS,0);   *          INITIALIZE;
  ITER=ITER+1;                       *COUNT THE NUMBER OF ITERATIONS;
  DO I=1 TO NITEMS;                  *BEGIN THE ITEM CALIBRATION LOOP;
  REPEATD: DO G=1 TO NITEMS-1;                *     FOR EACH SCORE GROUP;
           E=(EXP(THETA'(,G)-B(,I)));*COMPUTE THE NUMERATOR OF PROB;
           RPROB(G,I)=E#/(1+E); END;   *COMPUTE THE ITEM RASCH PROB;
  SUMNPQ=SUM(RPROB(,I)#(1-RPROB(,I))#N); *SUM NPQ OVER SCORE GROUPS;
  SUMNP=SUM(RPROB(,I)#N);                * SUM NP OVER SCORE GROUPS;
  CHANGE=(ISCORE(,I)-SUMNP)#/SUMNPQ;     *  COMPUTE CHANGE ITEM CAL;
```

```
 B(,I)=B(,I)-CHANGE;                          *  MODIFY ITEM CALIBRATION;
 IF ABS(CHANGE)  = 0.01 THEN GOTO REPEATD; *IF SMALL CHANGE, GO ON;
 SUMD2=SUMD2+(CHANGE#CHANGE); END;       *SUM SQUARED CHANGE VALUES;
 MEAND=SUM(B)#/NITEMS;                  *COMPUTE MEAN OF ITEM VALUES;
 B=(B-MEAND);                           *RECENTER ITEM CALIBRATIONS;
 DO I=1 TO NITEMS-1;                    *BEGIN CALIBRATION FOR     ;
 REPEATH: DO G=1 TO NITEMS;             *     PERSONS FOR EACH ITEM;
          E=(EXP(THETA1(,I)-B(,G)));    *     COMPUTE THE NUMERATOR;
          RPROB(I,G)=E#/(E+1); END;     *    COMPUTE THE RASCH PROB;
 SUMP=SUM(RPROB(I,));                   *  SUM THE PROBS OVER ITEMS;
 SUMPQ=SUM(RPROB(I,)#(1+RPROB(I,)));    *     SUM P(1+P) OVER ITEMS;
 CHANGE=(SUMP-I)#/SUMPQ;                *  CHANGE FOR PERSON VALUES;
 THETA1(,I)=THETA1(,I)-CHANGE;          *    SCORE GROUP RASCH VALUE;
IF ABS(CHANGE)  =0.01 THEN GOTO REPEATH; *  IF CHANGE SMALL, GO ON;
 END;                                  *END OF CALIBRATION ITERATION;
 IF SUMD2#/NITEMS  0.00001 THEN GOTO ITER8; *RECALIBRATE IF NEEDED;
 NOTE 'NUMBER OF ITERATIONS'; PRINT ITER;
 B=B#(NITEMS-1)#/NITEMS;          *ADJUST ITEM CALIBRATIONS FOR BIAS;
 NOTE 'FINAL ITEM CALIBRATIONS'; PRINT B;
 DO I=1 TO NITEMS;
 SED(,I)=1#/(SQRT(SUM(N#RPROB(,I)#(1-RPROB(,I))))); END;
 NOTE 'STANDARD ERROR FOR ITEM CALIBRATIONS';PRINT SED;
 NOTE 'FINAL PERSON CALIBRATIONS';PRINT THETA1;
 DO I=1 TO NITEMS-1;
 SEB(,I)=1#/(SQRT(SUM(RPROB(I,)#(1-RPROB(I,))))); END;
 NOTE 'STANDARD ERROR FOR PERSON CALIBRATIONS'; PRINT SEB;
                              *END OF RASCH CALIBRATION ROUTINE;
*****************************************************************
*   THIS SECTION OF THE PROGRAM COMPUTES THE MEAN SQUARE FIT    *
*               STATISTIC FOR  EACH  PERSON                     *
*             SEE WRIGHT AND STONE, BEST TEST DESIGN            *
*****************************************************************;
MSFIT=J.(NPERSONS,1,0);
DO I=1 TO NPERSONS;
   ZN=Y1J(I,)-RPROB(PSCORE(I,),);
   ZD=SQRT(RPROB(PSCORE(I,),)#(1-(RPROB(PSCORE(I,),))));
   ZI=ZN#/ZD;
   MSFIT(I,)=SSQ(ZI)#/(NITEMS-1);
END;
SUMFIT=SUM(MSFIT); NOTE 'SUM OF PERSON FIT STATISTICS'; PRINT SUMFIT;
*****************************************************************
*     THIS SECTION OF THE PROGRAM COMPUTES THE MEAN SQUARE FIT   *
*                  STATISTIC FOR EACH ITEM                      *
*       SEE WRIGHT AND STONE, BEST TEST DESIGN                  *
*****************************************************************;
ITEMFIT=J.(1,NITEMS,0);
DO I=1 TO NITEMS; Z1=0;
DO K=1 TO NPERSONS; RP=RPROB(PSCORE(K,),I);
ZT=Y1J(K,1)-RP;
ZB=RP#(1-RP);
Z1=ZT#ZT#/ZB+Z1; END;
ITEMFIT(,I)=Z1#/(NPERSONS-1); END;
NOTE 'ITEM MEAN SQUARE FIT VALUES'; PRINT ITEMFIT;
* REDEFINE VARIABLE NAMES FOR THE REMAINDER OF THE PROGRAM;
```

```
PIJ=J.(NPERSONS,NITEMS,0);
THETA=J.(NPERSONS,1,0);
DO I=1 TO NPERSONS;
THETA(I,)=THETA1(,PSCORE(I,));
PIJ(I,)=RPROB(PSCORE(I.)',1:NITEMS);END;
THETA=THETA';
   LINK INITIAL;      * INITIAL DETERMINES NEEDED VALUES;
   LINK ECIZ2;        * PROGRAM THEN LINKS ALL ECI CALCULATIONS;
   LINK ECIZ4;        * WHICH FUNCTION LIKE SUBROUTINES;
   STOP;

INITIAL:
*--------------STANDARDIZED EXTENDED CAUTION INDICES-------------*
*                                                                *
*                        ECIZ2 AND ECIZ4                         *
*                                                                *
*         AS DESCRIBED BY TATSUOKA & LINN, 1983                  *
*                        AND TATSUOKA, 1984                      *
*----------------------------------------------------------------*;
NROWYIJ=NROW(YIJ);
NCOLYIJ=NCOL(YIJ);
FREE B;
*------------------------INITIALIZATION--------------------------*
*     NEEDED VALUES OF YIJ ARE ELEMENTS OF YIJ                   *
*     NEEDED VALUES OF PIJ ARE ELEMENTS OF PIJ                   *
*                            TI=ITH ROW MEAN OF PIJ              *
*                            GJ=GTH COL MEAN OF PIJ              *
*                            G=MEAN OF GJ'S                      *
*                            VIJ=VARIANCE(PQ'S) OF PIJ           *
*                            GJG=VECTOR GJ-G                     *
*                            GJG2=SQUARE OF GJG                  *
*----------------------------------------------------------------*;
NCOLPIJ=NCOL(PIJ);          *GET NUMBER OF COLUMNS IN PIJ        ;
NROWPIJ=NROW(PIJ);          *GET NUMBER OF ROWS IN PIJ           ;
PIDOT=PIJ(,+);              *GET ROW SUM FOR PIJ                 ;
PDOTJ=PIJ(+,);              *GET COL SUM FOR PIJ                 ;
TI=PIDOT#/NCOLPIJ;          *COMPUTE ROW MEANS OF PIJ            ;
GJ=PDOTJ#/NROWPIJ;          *COMPUTE COL MEANS OF PIJ            ;
SUMGJ=GJ(,+);               *COMPUTE SUM OF GJS                  ;
G=SUMGJ#/NCOLPIJ;           *COMPUTE GRAND MEAN OF PIJ           ;
QMATRIX=1-PIJ;              *CREATE MATRIX OF QS (1-P)S          ;
VIJ=QMATRIX#PIJ;            *COMPUTE ELEMENT VARIANCES           ;
GJG=GJ-G;                   *COMPUTE VECTOR GJS-G                ;

RETURN;

ECIZ2:

*----------------ECIZ2 COMPUTATION FORMULA-----------------------*
*                                                                *
*             AS IN TATSUOKA, 1984, PSYCHOMETRICA, P. 105        *
*                                                                *
*                      N                    N                    *
*                   SUM PIJ(GJ-G)        SUM YIJ(GJ-G)           *
```

```
*            ECIZ2 = --------------   -   ---------------          *
*                      N           2 1/2     N           2 1/2     *
*                   (SUM VIJ(GJ-G) )       (SUM VIJ(GJ-G) )       *
*                                                                 *
*-----------------------------------------------------------------*;
PIJGJG=PIJ*GJG';            * STEP 1, FIRST COMPONENT, NUMERATOR     ;
GJG2=GJG#GJG;               * SQUARE GJG VECTOR                      ;
ONES=J(NROW PIJ,1,1);       * CREATE VECTOR OF ONES                  ;
MGJG2=ONES*GJG2;            * CREATE MATRIX OF REPEATING ROWS        ;
VIJGJG2=VIJ#MGJG2;          * STEP 1, FIRST COMPONENT, DENOMINATOR   ;
SVIJGJG2=VIJGJG2(,+);       * STEP 2, FIRST COMPONENT, DENOMINATOR   ;
B=SQRT(SVIJGJG2);           * STEP 3, FIRST COMPONENT, DENOMINATOR   ;
YIJGJG=YIJ*GJG';            * STEP 1, SECOND COMPONENT, NUMERATOR    ;
ECIZ2=(PIJGJG-YIJGJG)#/B;   * ECIZ2                                  ;
ECI2=PIJGJG-YIJGJG;
RETURN;                     * RETURN TO LINKS                        ;
ECIZ4:

*----------------ECIZ4 COMPUTATION FORMULA------------------------*
*                                                                 *
*            AS IN TATSUOKA, 1984, PSYCHOMETRICA, P. 105          *
*                                                                 *
*                     N                      N                    *
*                   SUM PIJ(PIJ-TI)        SUM YIJ(PIJ-TI)       *
*            ECIZ4 = --------------   -   ----------------        *
*                   N             2 1/2   N              2 1/2   *
*                (SUM VIJ(PIJ-TI) )     (SUM VIJ(PIJ-TI) )      *
*                                                                 *
*-----------------------------------------------------------------*;
ONES2=J(1,NCOLPIJ,1);
MTI=TI*ONES2;
PIJTI=PIJ-MTI;              * STEP 1, FIRST COMPONENT, NUMERATOR     ;
PPTI=PIJ#PIJTI;             * STEP 2, FIRST COMPONENT, NUMERATOR     ;
K=PPTI(,+);                 * STEP 3, FIRST COMPONENT, NUMERATOR     ;
PIJTI2=PIJTI#PIJTI;         * STEP 1, FIRST COMPONENT, DENOMINATOR   ;
VIJPIJT2=VIJ#PIJTI2;        * STEP 2, FIRST COMPONENT, DENOMINATOR   ;
L=VIJPIJT2(,+);             * STEP 3, FIRST COMPONENT, DENOMINATOR   ;
SQRTL=SQRT(L);
YIJTI=YIJ#PIJTI;            * STEP 1, SECOND COMPONENT, NUMERATOR    ;
M=YIJTI(,+);                * STEP 2, SECOND COMPONENT, NUMERATOR    ;
ECI4=K-M;
ECIZ4=ECI4#/SQRTL;          * ECIZ4                                  ;
THETA=THETA';
NOTE 'ID THETA ECI2 ECIZ2 ECI4 ECIZ4 MSFIT';
INDICES=NAME  THETA  ECI2  ECIZ2  ECI4  ECIZ4  MSFIT;
OUTPUT INDICES OUT=INDICES; * OUTPUT DATA SET ECIZ                   ;
PRINT INDICES;
RETURN;                     * GO BACK TO LINKS, THEN STOP            ;
NOCONV: NOTE 'DID NOT CONVERGE AFTER FIFTY ITERATIONS'; STOP;
DATA C;SET INDICES;FILE XYZ;
          PUT ROW COL1 COL2 COL3 COL4 COL5 COL6 COL7;
```

Table 1

Means and Standard Deviations of the Major Variables for All Available Examinees and for Examinees Used in Subsequent Analyses

Group	Variable	M	SD	n
	Course Performance	-0.159	1.187	557
	Algebra Ability	-0.383	1.41	706
	Trigonometry Ability	-0.734	0.876	531
	Algebra $ECI4_z$	-0.073	1.209	708
	Trigonometry $ECI4_z$	-0.008	1.073	531
Precalculus				
	Course Performance	-0.120	1.032	272
	Algebra Ability	-0.570	0.813	291
	Trigonometry Ability	-1.060	0.625	223
	Algebra $ECI4_z$	-0.193	1.194	292
	Trigonometry $ECI4_z$	0.115	1.144	223
Calculus				
	Course Performance	-0.052	1.072	148
	Algebra Ability	0.504	0.837	164
	Trigonometry Ability	-0.274	0.722	157
	Algebra $ECI4_z$	-0.341	1.063	165
	Trigonometry $ECI4_z$	-0.292	0.940	157

Note. Algebra ability equals the Rasch ability estimate derived from the algebra examinations. Trigonometry ability equals the Rasch ability estimate from the trigonometry examination. Course performance was the course section standardized semester point total. The ECI_zs were computed as described by Tatsuoka (1984, p. 105).

Table 2

Zero-order Correlations of the Major Variables for All Available Students (Elements Presented as r / Number of Observations)

Variable		(2)	(3)	(4)	(5)
Course Performance	(1)	.319	.272	-.114	-.152
		550	403	552	403
Algebra Ability	(2)		.680	-.091	-.281
			509	706	509
Trigonometry Ability	(3)			.004	-.156
				510	531
Algebra $ECI4_z$	(4)				.140
					510
Trigonometry $ECI4_z$	(5)				

Note. Algebra and trigonometry abilities were equal to the Rasch ability estimate from the respective examinations. Course performance was the course section standardized semester point total. The $ECI4_z$ were computed as described by Tatsuoka (1984, p. 105).

Table 3

Zero-order Correlations of the Major Variables for Precalculus Students (Elements Presented as r / Number of Observations)

Variable		(2)	(3)	(4)	(5)
Course Performance	(1)	.354	.279	-.101	-.227
		270	209	271	209
Algebra Ability	(2)		.356	-.024	-.208
			223	291	223
Trigonometry Ability	(3)			-.016	-.056
				223	223
Algebra $ECI4_z$	(4)				.191
					223
Trigonometry $ECI4_z$	(5)				

Note. Algebra and trigonometry abilities were equal to the Rasch ability estimate from the respective examinations. Course performance was the course section standardized semester point total. The $ECI4_z$ were computed as described by Tatsuoka (1984, p. 105).

Table 4

Zero-order Correlations of the Major Variables for Calculus Students (Elements Presented as r / Number of Observations)

Variable		(2)	(3)	(4)	(5)
Course Performance	(1)	.487	.454	.054	-.107
		143	138	144	138
Algebra Ability	(2)		.591	.153	-.174
			156	164	156
Trigonometry Ability	(3)			.142	-.063
				157	157
Algebra $ECI4_z$	(4)				.017
					157
Trigonometry $ECI4_z$	(5)				

Note. Algebra and trigonometry abilities were equal to the Rasch ability estimate from the respective examinations. Course performance was the course section standardized semester point total. The $ECI4_z$ were computed as described by Tatsuoka (1984, p. 105).

Table 5

Sequential Sums of Squares of Variables for Precalculus Students

Order	Variable		Type 1 SS	F	p<F	r^2
1	Algebra Ability	(A)	23.698	26.92	.0001	.107
2	Trigonometry Ability	(B)	6.793	7.72	.0060	.138
3	Algebra $ECI4_z$	(C)	4.529	5.14	.0244	.159
4	Trigonometry $ECI4_z$	(D)	4.780	5.43	.0208	.180
5	A * C		1.269	1.44	.2313	.186
6	B * D		1.721	1.95	.1637	.194

Note. Algebra ability equals the Rasch ability estimate derived from the algebra examinations. Trigonometry ability equals the Rasch ability estimate from the trigonometry examination. Course performance was the course section standardized semester point total.

The ECI_zs were computed as described by Tatsuoka (1984, p. 105). The r^2 values were computed at each sequential step to provide an indication of the increment to variance accounted for by the inclusion of each new variable. Values were based on regression analysis predicting course performance (n=209)

Table 6

Sequential Sums of Squares of Variables for Calculus Students

Order	Variable		Type 1 SS	F	p<F	r^2
1	Algebra Ability	(A)	37.363	45.46	.0001	.258
2	Trigonometry Ability	(B)	6.395	7.79	.0061	.292
3	Algebra $ECI4_z$	(C)	.205	.25	.6183	.294
4	Trigonometry $ECI4_z$	(D)	.012	.20	.9023	.294
5	A * C		.368	.45	.5046	.296
6	B * D		.243	.30	.5880	.297

Note. Algebra ability equals the Rasch ability estimate derived from the algebra examinations. Trigonometry ability equals the Rasch ability estimate from the trigonometry examination. Course performance was the course section standardized semester point total. The ECI_zs were computed as described by Tatsuoka (1984, p. 105). The r^2 values were computed at each sequential step to provide an indication of the increment to variance accounted for by the inclusion of each new variable. Values were based on regression analysis predicting course performance (n=137).

Table 7

Mean Comparisons of Aberrantly vs. Consistently Responding Groups by Course Placement

Variable	Placement	F	p<F
Groups Based on Algebra or Trigonometry Responses			
Algebra Ability	Precalculus	0.06	.8084
	Calculus	5.02	.0265[1]
Trigonometry Ability	Precalculus	0.05	.8180
	Calculus	1.56	.2137
Groups Based on Algebra Responses			
Course Performance	Precalculus	1.30	.2557
	Calculus	1.06	.3053
Groups Based on Trigonometry Responses			
Course Performance	Precalculus	7.10	.0083[2]
	Calculus	.06	.4245

Note. Associated sample sizes were 270 precalculus algebra responses, 209 precalculus trigonometry responses, 143 calculus algebra responses, and 138 calculus trigonometry responses. Groups were formed by placing students with $ECI4_z$s in the top one-third of the distribution one group and all others in the second.

[1] Aberrant Group (mean=0.746, n=38)
Consistent Group (mean=0.417, n=105)

[2] Aberrant Group (mean=-0.283, n=85)
Consistent Group (mean= 0.098, n=124)

Table 8

Relationship Between Mathematical Ability and Course Performance by Aberrance Group and by Course Placement

Ability	Course Placement			
	Precalculus		Calculus	
	Aberrant	Consistent	Aberrant	Consistent
Algebra (n)	77	193	38	105
Beta	.507	.437	.624	.619
SE	.193	.072	.155	.120
Intercept	.036	.181	- .382	- .381
Trigonometry (n)	85	124	23	115
Beta	.673	.241	.652	.666
SE	.145	.152	.235	.131
Intercept	.430	.356	.048	.115

Note. Students with $ECI4_z$s in the top one-third of the distribution were placed into the aberrant group. All others were placed into the consistent group.

Table 9

Examination of Predictive Bias by Course Placement: Regression Slope Differences

Course Placement	Ability	t	df	p<t
Precalculus	Algebra	.340	266	.7343
	Trigonometry	2.031	205	.0436
Calculus	Algebra	.0252	139	.9800
	Trigonometry	.0535	134	.9574

Note. Results were based on the regression of algebra or trigonometry on course performance.

Table 10

Examination of Predictive Bias by Course Placement: Standard Error of Estimate Differences

Course Placement	Ability	F	$df(n_1-2/n_2-2)$	p<F
Precalculus	Algebra	2.639	75/191	.0001
	Trigonometry	1.027	83/122	.4423
Calculus	Algebra	1.2917	36/103	.1602
	Trigonometry	1.7938	21/113	.0273

Note. Values were based on F test of the ratio of standard errors of estimates from regression of math ability on course grade.

Table 11

Means and Standard Deviations of the Major Variables for All Available Examinees and for Examinees Used in Subsequent Analyses for the Replication

Group	Variable	M	SD	n
	Grade	2.411	1.321	4579
	Algebra Ability	-0.354	.988	4339
	Trigonometry Ability	-0.874	0.758	3563
	Algebra $ECI4_z$	-0.078	1.180	4339
	Trigonometry $ECI4_z$	-0.009	1.120	3563
Precalculus				
	Grade	2.452	1.304	2486
	Algebra Ability	-0.524	0.761	2402
	Trigonometry Ability	-1.067	0.603	2055
	Algebra $ECI4_z$	-0.082	1.186	2402
	Trigonometry $ECI4_z$	0.091	1.134	2055
Calculus				
	Grade	2.524	1.348	955
	Algebra Ability	0.533	0.829	888
	Trigonometry Ability	-0.360	0.670	875
	Algebra $ECI4_z$	-0.344	1.071	888
	Trigonometry $ECI4_z$	-0.362	1.010	875

Note. Algebra ability equals the Rasch ability estimate derived from the algebra examinations. Trigonometry ability equals the Rasch ability estimate from the trigonometry examination. Course performance was the course section standardized semester point total. The ECI_zs were computed as described by Tatsuoka (1984, p. 105).

Table 12

Zero-order Correlations of the Major Variables for All Available Students in Replication (Elements Presented as r / Number of Observations)

Variable		(2)	(3)	(4)	(5)
Grade	(1)	.410	.253	-.145	-.134
		4201	3465	4201	3465
Algebra Ability	(2)		.622	-.157	-.240
			3563	4339	3563
Trigonometry Ability	(3)			-.082	-.193
				3563	3563
Algebra $ECI4_z$	(4)				.164
					3563
Trigonometry $ECI4_z$	(5)				

Note. Algebra and trigonometry abilities were equal to the Rasch ability estimate from the respective examinations. Course performance was the course section standardized semester point total. The $ECI4_z$ were computed as described by Tatsuoka (1984, p. 105).

Table 13

Zero-order Correlations of the Major Variables for Precalculus Students in Replication (Elements Presented as r / Number of Observations)

Variable		(2)	(3)	(4)	(5)
Grade	(1)	.370	.173	-.161	-.121
		2321	1991	2321	1991
Algebra Ability	(2)		.378	-.155	-.196
			2055	2402	2055
Trigonometry Ability	(3)			-.076	-.129
				2055	2055
Algebra $ECI4_z$	(4)				.157
					2055
Trigonometry $ECI4_z$	(5)				

Note. Algebra and trigonometry abilities were equal to the Rasch ability estimate from the respective examinations. Course performance was the course section standardized semester point total. The $ECI4_z$ were computed as described by Tatsuoka (1984, p. 105).

Table 14

Zero-order Correlations of the Major Variables for Calculus Students in Replication (Elements Presented as r / Number of Observations)

Variable		(2)	(3)	(4)	(5)
Grade	(1)	.493	.364	-.009	-.091
		869	856	869	856
Algebra Ability	(2)		.589	.024	-.118
			875	888	875
Trigonometry Ability	(3)			.054	-.086
				875	875
Algebra $ECI4_z$	(4)				.045
					875
Trigonometry $ECI4_z$	(5)				

Note. Algebra and trigonometry abilities were equal to the Rasch ability estimate from the respective examinations. Course performance was the course section standardized semester point total. The $ECI4_z$ were computed as described by Tatsuoka (1984, p. 105).

Table 15

Replication Sequential Sums of Squares of Variables for Precalculus Students

Order	Variable		Type 1 SS	F	p<F	r^2
1	Algebra Ability	(A)	421.343	299.15	.0001	.129
2	Trigonometry Ability	(B)	5.070	3.60	.0579	.131
3	Algebra $ECI4_z$	(C)	30.818	21.88	.0001	.140
4	Trigonometry $ECI4_z$	(D)	4.518	3.21	.0734	.142
5	A * C		.155	.11	.7401	.142
6	B * D		.424	.30	.5834	.142

Note. Algebra ability equals the Rasch ability estimate derived from the algebra examinations. Trigonometry ability equals the Rasch ability estimate from the trigonometry examination. Course performance was the course section standardized semester point total.

The ECI_zs were computed as described by Tatsuoka (1984, p. 105). The r^2 values were computed at each sequential step to provide an indication of the increment to variance accounted for by the inclusion of each new variable. Values were based on regression analysis predicting course performance (n=209).

Table 16

Replication Sequential Sums of Squares of Variables for Calculus Students

Order	Variable		Type I SS	F	p<F	r^2
1	Algebra Ability	(A)	356.627	272.86	.0001	.240
2	Trigonometry Ability	(B)	13.396	10.25	.0014	.249
3	Algebra $\overline{ECI4}_z$	(C)	.932	.71	.3987	.250
4	Trigonometry $\overline{ECI4}_z$	(D)	1.642	1.26	.2627	.251
5	A * C		2.014	1.54	.2148	.252
6	B * D		1.518	1.16	.2815	.253

Note. Algebra ability equals the Rasch ability estimate derived from the algebra examinations. Trigonometry ability equals the Rasch ability estimate from the trigonometry examination. Course performance was the course section standardized semester point total. The $\overline{ECI}_z$s were computed as described by Tatsuoka (1984, p. 105). The r^2 values were computed at each sequential step to provide an indication of the increment to variance accounted for by the inclusion of each new variable. Values were based on regression analysis predicting course performance (n=137).

Table 17

Replication Mean Comparisons of Aberrantly vs. Consistently Responding Groups by Course Placement

Variable	Placement	F	p<F
Groups Based on Algebra or Trigonometry Responses			
Algebra Ability	Precalculus	14.40	.0002[1]
	Calculus	0.02	.8934
Trigonometry Ability	Precalculus	21.38	.0001[2]
	Calculus	0.68	.4107
Groups Based on Algebra Responses			
Grade	Precalculus	18.81	.0001[3]
	Calculus	.08	.7747
Groups Based on Trigonometry Responses			
Grade	Precalculus	14.94	.0001[4]
	Calculus	1.29	.2564

Note. Associated sample sizes were 270 precalculus algebra responses, 209 precalculus trigonometry responses, 143 calculus algebra responses, and 138 calculus trigonometry responses. Groups were formed by placing students with $ECI4_z$s in the top one-third of the distribution one group and all others in the second.

[1] Aberrant Group (mean= -.651, n=716)
Consistent Group (mean= -.454, n=1585)

[2] Aberrant Group (mean= -1.161, n=715)
Consistent Group (mean= -1.010, n=1263)

[3] Aberrant Group (mean= 2.233, n=716)
Consistent Group (mean= 2.592, n=1585)

[4] Aberrant Group (mean= 2.361, n=715)
Consistent Group (mean= 2.642, n=1263)

Table 18

Relationship Between Mathematical Ability and Course Performance by Aberrance Group and by Course Placement for Replication Groups

Ability	Course Placement			
	Precalculus		Calculus	
	Aberrant	Consistent	Aberrant	Consistent
Algebra (n)	716	1585	198	667
Beta	.630	.614	.667	.842
SE	.068	.038	.098	.054
Intercept	2.644	2.871	2.188	2.121
Trigonometry (n)	715	1263	176	680
Beta	.325	.358	.860	.667
SE	.084	.057	.119	.074
Intercept	2.738	3.004	2.812	2.853

Note. Students with $ECI4_z$s in the top one-third of the distribution were placed into the aberrant group. All others were placed into the consistent group.

Table 19

Examination of Predictive Bias by Course Placement: Regression Slope Differences for Replication Data

Course Placement	Ability	t	df	p<t
Precalculus	Algebra	.205	2297	.8376
	Trigonometry	.325	1974	.7452
Calculus	Algebra	1.564	861	.1182
	Trigonometry	1.379	852	.1683

Note. Results were based on the regression of algebra or trigonometry on course performance.

Table 20

Examination of Predictive Bias by Course Placement: Standard Error of Estimate Differences for Replication Data

Course Placement	Ability	*F*	$df(n_1-2/n_2-2)$	$p<F$
Precalculus	Algebra	1.125	714/1583	.0309
	Trigonometry	1.053	713/1261	.2156
Calculus	Algebra	1.285	196/665	.0122
	Trigonometry	1.175	174/678	.0829

Note. Values were based on *F* test of the ratio of standard errors of estimates from regression of math ability on course performance.